European Journal of Epidemiology
https://doi.org/10.1007/s10654-022-00908-y

REVIEW

The methodologies to assess the effectiveness of non-pharmaceutical interventions during COVID-19: a systematic review

Nicolas Banholzer[1] · Adrian Lison[2] · Dennis Özcelik[3] · Tanja Stadler[2] · Stefan Feuerriegel[1,4] · Werner Vach[5,6]

Received: 20 April 2022 / Accepted: 15 July 2022

Abstract

Non-pharmaceutical interventions, such as school closures and stay-at-home orders, have been implemented around the world to control the spread of SARS-CoV-2. Their effectiveness in improving health-related outcomes has been the subject of numerous empirical studies. However, these studies show fairly large variation among methodologies in use, reflecting the absence of an established methodological framework. On the one hand, variation in methodologies may be desirable to assess the robustness of results; on the other hand, a lack of common standards can impede comparability among studies. To establish a comprehensive overview over the methodologies in use, we conducted a systematic review of studies assessing the effectiveness of non-pharmaceutical interventions between January 1, 2020 and January 12, 2021 (n = 248). We identified substantial variation in methodologies with respect to study setting, outcome, intervention, methodological approach, and effectiveness assessment. On this basis, we point to shortcomings of existing studies and make recommendations for the design of future studies.

Keywords Non-pharmaceutical interventions · Social distancing measures · Control measures · COVID-19 · Systematic review · Methodology review

Introduction

In response to the COVID-19 pandemic, countries around the world have implemented non-pharmaceutical interventions. These include a variety of public health measures implemented by governments at the population level to decrease the number of person-to-person contacts with the intention of controlling, preventing, and mitigating transmission, e.g. school closures, stay-at-home orders, and mandates for compulsory wearing of masks in public places [1–4]. The widespread use of these interventions has raised interest in empirically studying their effects on health-related outcomes reflecting disease dynamics, e. g. the number of new cases or infection rates [5–10]. Such studies can play an important role in informing the discussion about the effectiveness of interventions. In particular, insights from the COVID-19 pandemic may contribute to an evidence-based public-health response in subsequent COVID-19 waves or future pandemics.

Accordingly, a plethora of studies assessing the effectiveness of non-pharmaceutical interventions during the COVID-19 pandemic have been published. Their findings have been summarized by several meta-analyses [11–15];

Nicolas Banholzer and Adrian Lison contributed equally to this work.

✉ Nicolas Banholzer
nbanholzer@ethz.ch

✉ Adrian Lison
adrian.lison@bsse.ethz.ch

1 Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

2 Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland

3 Chemistry | Biology | Pharmacy Information Center, ETH Zurich, Zurich, Switzerland

4 LMU Munich School of Management, LMU Munich, Munich, Germany

5 Basel Academy for Quality and Research in Medicine, Basel, Switzerland

6 Department of Environmental Sciences, University of Basel, Basel, Switzerland

Published online: 24 September 2022

nonetheless, each meta-analysis considered a different subset of studies. We argue that the latter is due to substantial variation in the methodologies used to conduct empirical studies on the effectiveness of non-pharmaceutical interventions. The resulting lack of similarity constrains meta-analyses to a comparably small and specific subset of the overall evidence.

There are different reasons to expect variation in methodologies in the studies on the effectiveness of non-pharmaceutical interventions for controlling a pandemic. One possibility is the lack of empirical data before the COVID-19 pandemic, so that early studies have been largely theoretical [16]. Empirically assessing the effectiveness of non-pharmaceutical interventions is therefore a relatively new subject, and corresponding studies do not build on an established scientific framework. Another possibility is that empirical assessments have been approached with different methods and domain knowledge by researchers from various fields, e. g. computational biology, infectious disease epidemiology, public health, economics, and statistical modeling.

Variation in methodologies can be manifold. Different study settings, outcomes, interventions, methodological approaches, and ways to assess effects may be used. On the one hand, such variation may be desired as it allows to assess the robustness of results against individual assumptions and methodologies. On the other hand, variation in methodologies can impede comparability among studies, which is necessary to arrive at conclusive evidence regarding the effectiveness of non-pharmaceutical interventions.

Here, we systematically review the methodologies for studying the associations of non-pharmaceutical interventions with health-related outcomes. Thereby, we aim to inform about different methodologies that were used by previous studies and identify opportunities to improve the robustness and comparability of future studies. In particular, we explore shortcomings of common approaches and provide recommendations for subsequent studies on the effectiveness of non-pharmaceutical interventions.

Methods

We tailored our review to the challenge of mapping a potentially diverse set methodologies from a large number of studies. To ensure rigour and consistency, we preregistered the procedures for all stages of the review process, following common guidelines for systematic literature reviews [17]. Certain guidelines were not applicable to a methodology review as ours. In particular, our eligibility criteria and risk of bias assessment reflect the objective of this review, which was not to evaluate the evidence regarding the effectiveness of non-pharmaceutical interventions, but to map the variation in methodologies used. The preregistered methodology was documented in a review protocol at PROSPERO [18].

We report our review according to the PRISMA 2020 statement [19]. A completed PRISMA 2020 checklist is provided in online Appendix G. The search strategy was developed jointly and executed by an experienced information consultant. Then, two authors (NB and AL) performed study selection, data extraction, and synthesis, while having regular meetings with the complete author team.

Eligibility criteria for studies

In the following, we describe our eligibility criteria, which informed our search strategy and were systematically applied during study selection. Importantly, if a study contained multiple analyses of which only some fulfilled our eligibility criteria, we included the study but extracted only the eligible analyses. This may sometimes not correspond to the main analysis of a paper or may include more than one analysis per study.

Study design

In this review, we considered observational studies assessing the associations of non-pharmaceutical interventions with outcomes related to the COVID-19 disease. We focused on analyses that used real-world observational data to assess the effectiveness of non-pharmaceutical interventions. Specifically, we excluded modeling studies that predominantly worked with synthetic data or projected future transmission dynamics based on hypothetical scenarios without assessing the effectiveness of interventions empirically.

We considered all types of observational study designs, i. e. cross-sectional, case-control, retrospective and prospective cohort, etc. However, note that these study designs often relate to the analysis of individuals or cohorts. This is different from our setting where non-pharmaceutical interventions were implemented at the population level, thus typically all individuals within a population were exposed to non-pharmaceutical interventions at the same time. As a result, changes in outcomes following interventions were also usually evaluated at the population rather than the individual level. Nevertheless, our review sample includes studies that used individual-level epidemiological data, e. g. individual-level epidemiological data on cases and transmission chains, which were typically used to compute population-level outcomes quantifying transmission such as the reproduction number.

Population

We considered studies assessing the effectiveness of non-pharmaceutical interventions on the population in one or several geographic regions. Our review was not limited to a specific geographic region, i. e. all national and subnational

regions worldwide were considered. We furthermore included studies analyzing specific subpopulations in a certain region (e. g. certain age groups). We also considered analyses using individual-level data, as long as the effectiveness of non-pharmaceutical interventions were assessed on a population level.

Outcome

The main outcomes considered were health-related outcomes at population level that are associated with COVID-19 (e. g. confirmed cases, hospitalizations, and deaths), and epidemiological outcomes characterizing infection dynamics such as reproduction numbers or transmission rates. We also considered similar outcomes associated with other infectious diseases (e. g. influenza), if used as a surrogate for COVID-19. Moreover, behavioral outcomes potentially mediating the effects of non-pharmaceutical interventions were included (e. g. human mobility). In contrast, we excluded analyses assessing the effects of non-pharmaceutical interventions solely on other outcomes not directly related to infectious diseases (e. g. psychological well-being or economic activities).

Intervention

As non-pharmaceutical interventions, we considered the implementation of health policy measures in the context of the COVID-19 pandemic. Specifically, we included any intervention related to social distancing (e. g. school closures, venue closures, workplace closures), containment (e. g. contact tracing, quarantining), population flow (e. g. border closures), or personal protection (e. g. facial mask mandates). Analyses were considered regardless of whether they assessed the effectiveness of a single intervention, the effectiveness of multiple interventions separately, or the effectiveness of a combination of interventions. For simplicity, we refer to these as non-pharmaceutical interventions throughout the review, while recognizing that also other terms have been used in the literature. Importantly, we accounted for various alternative terms in our literature search (see Search strategy below). We excluded interventions not directly related to disease control (e. g. economic measures like social benefits).

Search strategy

We searched for peer-reviewed original research articles in English language that were accepted, published, or in press between January 1, 2020 and January 12, 2021 (2929 unique records). In our review protocol, we specified that we would also include preprints in our search. However, due to their enormous volume, we eventually decided not to consider gray literature or preprints in our review. Our results therefore only cover methodologies used by articles peer-reviewed at the time of search, among which we already found considerable variation. To account for potentially new methodologies in articles published after the time of search, we also considered further recent studies on the effectiveness of non-pharmaceutical interventions in our discussion and put them into the context of our review findings.

We searched the databases Embase, PubMed, Scopus, and Web of Science. These databases include, among others, MEDLINE, Biological Abstracts, CAB Abstracts, and Global Health. We composed our search query of four components to be contained in the publication title or abstract: (1) a synonym for "non-pharmaceutical intervention", (2) a synonym for "estimation" or "assessment", (3) a synonym for "effect", and (4) a synonym for "COVID-19". Starting from a precompiled list of 18 references based on our primary research on the effectiveness of non-pharmaceutical interventions, we created and repeatedly extended a collection of synonyms for each of the above components, thereby achieving a broad search while keeping the number of selected studies manageable. The strings for our search queries are provided in online Appendix B. Importantly, we decided not to include search terms for single interventions such as face masks or travel restrictions, as this would have resulted in an unmanageable number of studies that were not concerned with the population-level impact of the non-pharmaceutical intervention. Nevertheless, our search query found studies on single interventions through other terms describing non-pharmaceutical interventions.

Data collection and analysis

Study selection

As a first step, we screened the titles of the studies retrieved from the database search for keywords clearly suggesting that the study would not meet our predefined eligibility criteria (e. g. "mental health" or "air quality"). The compiled set of keywords (see online Appendix B) was used to automatically identify cases for exclusion via the publication title. For all remaining studies, two authors (AL and NB) checked the eligibility criteria and individually decided on inclusion or exclusion. Each of the two authors checked the eligibility for one half of the studies via the following process: First, studies were checked by title and, if in doubt, by abstract. Then, if still in doubt, studies were checked by full text and discussed by both authors. Any disagreements were resolved with involvement of a third author (WV). Generally, we followed an inclusive approach by keeping all studies that could not be excluded with high confidence. At each stage, all decisions were recorded in a spreadsheet.

Data extraction

We extracted data from all included studies (n = 248) in a spreadsheet. Our extraction strategy reflected the exploratory nature of our analysis and thus allowed for new data items to be added throughout the process. Therefore, we maintained a detailed manual with all data items and the potential values for each item (see online Appendix D). Before extraction, a preliminary version of the data extraction form was created based on reporting items from checklists for observational studies (STROBE [20], RECORD [21]), a template for public health policy interventions (TIDieR-PHP [22]) and an initial set of studies. Aside from bibliographic information, the data to be extracted consisted of information on the study setting, outcome, intervention, methodological approach, and effectiveness assessment.

The extraction process was structured in four rounds. During the first round, two authors (AL and NB) extracted data from an initial set of 20 publications, blinded to each other's coding. The coding was then compared, and any differences were discussed to resolve ambiguities. Corresponding changes were recorded by updating the extraction form and manual, and applied subsequently. In the second round, the two reviewers each extracted data from one half of the remaining publications and checked the other half coded by their colleague. Color-coding was used to highlight uncertain or ambiguous entries for the other reviewer or to mark such entries for further discussion. Regular meetings were held between the two authors (AL and NB) to discuss these uncertainties and ambiguities. All disagreements were resolved through discussion, if needed by involving a third author (WV). Thereby, the data extraction manual and form were continuously refined and kept up-to-date. In particular, the list of values that could be potentially assigned to each data item was continuously extended and harmonized as new studies were extracted. In the third round, the data extraction form and manual were simplified by merging data items or categories that, retrospectively, were found redundant, or by relabeling items and categories to define them more precisely. This was done with particular attention to enable comparability among the extracted analyses as well as readability of the results. In the fourth round, the final scheme was applied to all studies.

Quality assessment

The goal of this systematic review was not to perform a meta-analysis or narrative synthesis of the evidence regarding the effectiveness of non-pharmaceutical interventions, but to compare the included studies along methodological dimensions and to analyze the variation in study setting, outcome, intervention, methodological approach, and effectiveness assessment. Therefore, no risk of bias assessment with regard to the study results was conducted. Our minimum requirement for quality was that most information on the aforementioned dimensions could be extracted from the manuscript and/or supplementary material. This minimum requirement was not met by four studies, which were thus excluded. For other studies where only some methodological information was missing, we noted in the data extraction sheet that this information "could not be evaluated".

Data synthesis

The results of the data extraction were synthesized in tabular form by recording the frequency of categories per item. We reported the frequency for each item of the main dimensions (study setting, outcome, intervention, methodological approach, and effectiveness assessment) individually. For some items, we conducted further specialized analyses, for example by computing the frequencies of categories for different methodologies separately, or by qualitatively evaluating the supplementary information added to certain entries during extraction. Insights from these additional analyses were reported textually. Furthermore, we synthesized common analysis types based on patterns identified in the methodological approaches. Lastly, based on our findings, we derived specific recommendations for future studies with regard to scope, robustness, and comparability, and put them into the context of more recent studies that were not part of our review sample.

Results

We conducted a systematic database search for peer-reviewed research articles from January 1, 2020 up to January 12, 2021 (see "Methods" section). Figure 1 shows the PRISMA flow diagram of our identification process. The search yielded 2,929 unique records of studies for screening. Through title and abstract screening, we identified 411 studies as potentially relevant and evaluated their full texts. Of these, we excluded 163 studies that did not meet the eligibility criteria. The most frequent reasons for exclusion were that (1) studies primarily simulated the effects of interventions in hypothetical scenarios rather than making inferences from observational data; (2) studies had a different objective than assessing the effectiveness of interventions, and (3) studies only assessed the association of health-related outcomes with population behavior (most often mobility), but not with interventions. The remaining n = 248 studies met our eligibility criteria and were included for subsequent data extraction. Importantly, 35 studies in our review sample contained multiple (i. e. up to three) analyses, e. g. with different methodological approaches, leading to 285 different

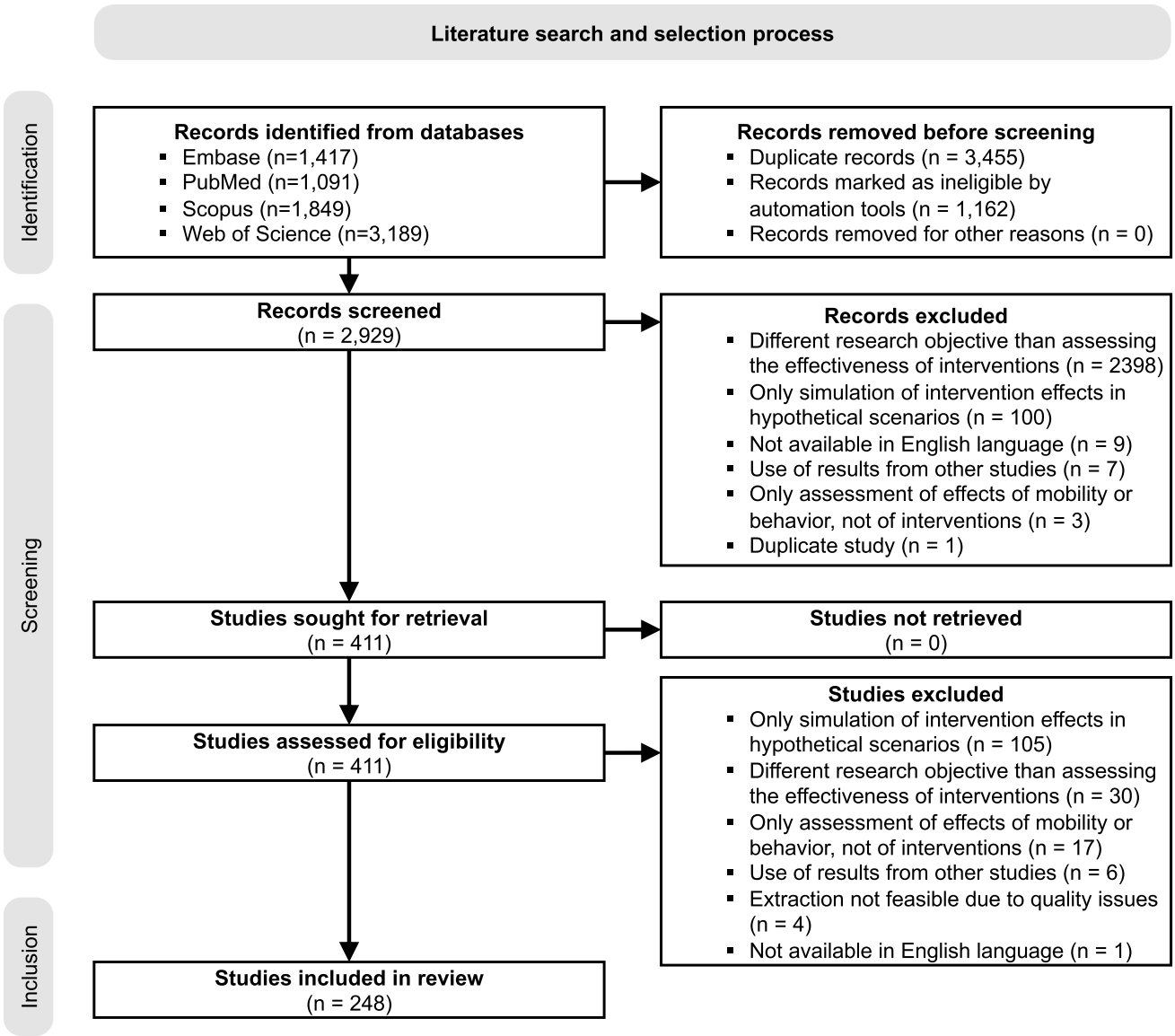

Fig. 1 PRISMA flow diagram. Overall, n = 248 studies were included. Some studies contain multiple analyses, such that the number of analyses included in the review is 285

analyses included. If not indicated otherwise, our results are presented at the level of individual analyses (and *not* at the level of studies).

We characterized the analyses along five dimensions (online Appendix D): study setting (D.1), outcome (D.2), intervention (D.3), methodological approach (D.4), and effectiveness assessment (D.5). In the Results section, if not stated otherwise, we use the term *interventions* to refer to non-pharmaceutical interventions. Where appropriate, we also point to exemplary studies of specific characteristics. Due to the large size of our review sample, however, we refrain from referencing all studies in the main manuscript and instead refer to our complete data extraction report in online Appendix E.

Study setting

The analyses vary in their scope across populations, geographic areas, and study period. A systematic classification of the study setting is shown in Table 1.

Population

More than half of the analyses studied multiple populations, i. e. multiple countries or subnational regions (e. g. states or cities). The remainder focused on a single population, i. e. a single country or subnational region. The analyses were performed at the national level, the subnational level, or both. If both levels were studied, the country and all its subnational

Table 1 Systematic classification and frequency of the study setting (D.1)

D.1.1: Number of populations included	Frequency
Single (country, state, city, etc.)	118 (41%)
Multiple (countries, states, cities, etc.)	167 (59%)
D.1.2: Level of populations included	
National (country-level)	117 (41%)
Subnational (e. g. state-level)	71 (25%)
Both national and subnational (country- and e. g. state-level)	97 (34%)
D.1.3: Geographic areas covered‡	
Asia	144 (51%)
Europe	109 (38%)
North America	91 (32%)
Middle East and Africa	49 (17%)
Central and South America	46 (16%)
Oceania	42 (15%)
D.1.4: Number of countries covered	
Multiple countries	66 (23%)
Single country (including multiple populations from a single country)‡	219 (77%)
China	54 (25%)
United States	43 (20%)
India	11 (5%)
Italy	11 (5%)
Other	100 (46%)
D.1.5: Study period	
Start and end date span first epidemic wave	161 (56%)
One or more exceptions‡	124 (44%)
End date in growth phase of wave	44 (35%)
Same end date for several populations with diverse epidemic trajectories	38 (31%)
End date at peak of wave	16 (13%)
End date could not be evaluated	14 (11%)
Other	14 (11%)

‡ Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100%

regions were oftentimes considered, e. g. all states of the United States. Geographically, some regions and countries were more frequently studied than others, probably due to an earlier start of the pandemic or particularly high incidence and mortality during the first epidemic wave.

Study period

Typically, the study period covered both a rise and decline in new cases of the first epidemic wave in the analyzed population, and started before and ended after the analyzed interventions were implemented. However, many analyses also deviated from this pattern in one or several aspects (Table 1, D.1.5). Notably, in some analyses, the end date of the study period was still within the epidemic growth phase for some populations but already in the control phase for other populations.

Outcome

The studies in our review sample used different types of health-related outcomes or surrogates. For every analysis, we identified the "raw outcome", i. e. the outcome data which were self-collected or obtained from external sources and used as input for the analysis. In around half of the analyses, the raw outcome was analyzed directly to assess the effectiveness of interventions. The other half of analyses, however, involved an intermediate step, in which another outcome was computed from the raw outcome. This

"computed outcome" was then analyzed instead of the raw outcome, or sometimes in addition to it. A systematic classification of the outcomes are shown in Table 2.

Raw outcome

We identified three main types of raw outcome data used, namely (1) epidemiological population-level data, (1) epidemiological individual-level data, and (3) behavioral data.

(1) *Epidemiological population-level data*

The majority of analyses used population-level data on epidemiological outcomes, e. g. confirmed cases and deaths. The most frequent types were surveillance data, mainly the number of confirmed cases, but also deaths, hospitalizations, recovered cases, and, less frequently, intensive care unit (ICU) admissions. Importantly, some outcomes, such as recovered cases, were predominantly used to fit transmission models, in which case the effectiveness of interventions was rather measured in terms of a different latent outcome (see D.5 Methodological approach, Sect. 3.4). Frequently, authors also included several types of data (e. g. both cases and deaths), either to perform a separate analysis for each (e. g. as a robustness check) or to combine them in a joint model (e. g. a transmission model). Some analyses used surveillance data on other diseases than COVID-19, with influenza being the most popular choice. Such surrogate diseases have often been monitored over an extended period of time, which allows comparing their spread during the COVID-19 pandemic to earlier years. Notably, we found only three analyses that used external data on latent epidemiological population-level outcomes (e. g. the reproduction number). All other analyses using a latent outcome self-computed it from raw data in an intermediate step (see D.2.2 Computed outcome, Sect. 3.2.2).

(2) *Epidemiological individual-level data*

Instead of population-level data, some analyses also used individual-level epidemiological data. These were in particular data about individual cases with case ID, demographics, and epidemiological characteristics (e. g. the date of symptom onset or travel history). In some instances, this included contact tracing data with links between index and secondary cases, allowing the reconstruction of transmission chains. Two analyses also used genome sequence data of clinical SARS-CoV-2 samples [23, 24].

(3) *Behavioral data*

In addition to epidemiological data, a relevant share of analyses employed data on population behavior, mainly mobility data. These data were usually obtained through tracking of mobile phone movements and provided as aggregates at the population level, based on summary statistics such as the daily number of trips made, time spent at certain locations, or population flow between regions. Another, less frequently used source of information on human behavior were surveys regarding social distancing practices, such as adherence to interventions, face mask usage, or daily face-to-face contacts.

Computed outcome

Around half of the analyses involved an intermediate step in which the raw outcome was used to compute another outcome, and the effectiveness of interventions were subsequently assessed using this "computed outcome". Typically, the rationale of such analyses was to conduct the assessment on an outcome with clearer epidemiological interpretation and relevance to policy-making. At the same time, the methodological separation of outcome computation and effectiveness assessment into distinct steps allowed the authors to limit the complexity of their models. In our review sample, we identified four main types of computed outcomes:

(1) Measures of epidemic trend were computed to describe the overall trend of the epidemic, e. g. through the growth rate or doubling time of confirmed cases or hospitalizations. These measures were often interpreted as crude estimates of the infection dynamics in a population, and authors used them to achieve better comparability of outcomes across different populations.
(2) Epidemiological parameters were computed to measure specific infection dynamics, most often in terms of the reproduction number. That is, studies typically used the time series of confirmed cases in a population to compute the effective reproduction number over time and then assessed whether it decreased during interventions. A few analyses also used individual-level epidemiological data to compute and assess changes in epidemiologically relevant time spans such as the serial interval or the time from symptom onset to isolation.
(3) Summary statistics were typically used to aggregate the raw outcome of a population over time into a single figure describing the progression of the epidemic in the population. For example, authors computed the time until a certain number of documented cumulative cases was reached, or the time until the reproduction number first fell below one.
(4) Change points in the outcome were computed with the aim to find time points of presumably structural changes in epidemic dynamics and compare them with implementation dates of interventions in the subsequent analysis [10, 25, 26]. Typically, change points were computed for the time series of confirmed cases or mobility.

Table 2 Systematic classification and frequency of the outcome (D.2)

D.2.1: Raw outcome‡	Frequency
Epidemiological population-level outcome‡	223 (78%)
Confirmed cases	186 (83%)
Deaths	64 (29%)
Recovered cases	20 (9%)
Hospitalizations	18 (8%)
Surrogate disease outcome	10 (4%)
Other	24 (11%)
Epidemiological individual-level outcome‡	23 (8%)
Individual cases	11 (48%)
Individual cases and transmission chains	8 (35%)
Genome sequence data	4 (17%)
Behavioral outcome‡	55 (19%)
Mobility	50 (91%)
Survey responses	6 (11%)
D.2.2: Time resolution of raw outcome	
Daily	269 (94%)
Other (weekly, biweekly, monthly, or not applicable)	16 (6%)
D.2.3: Computed outcome‡	
None (only raw outcomes used)	150 (53%)
Measure of epidemic trend‡	34 (12%)
Growth rate	24 (71%)
Doubling time	11 (32%)
Other	1 (3%)
Epidemiological parameter‡	89 (31%)
Reproduction number	78 (88%)
Transmission rate	6 (7%)
Other	16 (18%)
Summary statistic	8 (3%)
Change points	7 (2%)
Other	9 (3%)
D.2.4: Method to obtain the computed outcome	
None (no computed outcome)	150 (53%)
One or several methods used‡	135 (47%)
Simple computation (e. g. ratio, sum etc.)	35 (26%)
Exponential growth model	11 (8%)
Compartmental transmission model	35 (26%)
Statistical estimation of reproduction number	43 (32%)
Other	29 (21%)
D.2.5: Data source‡	
Could not be evaluated	10 (4%)
Data from (sub)national authorities	141 (49%)
Data from publicly available cross-country selections	77 (27%)
Mobility data from corporate organizations	40 (14%)
Other	54(19%)
D.2.6: Data availability	
Data access via source	173 (61%)
Data made available by the authors	76 (27%)
Data not accessible	36 (13%)

‡ Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100%

Of note, the raw outcome was not always used only for obtaining the computed outcome, e. g. changes both in the number of new confirmed cases (raw outcome) and in the reproduction number (computed outcome) were sometimes analyzed.

Method to obtain the computed outcome

(1) Measures of epidemic trend were often obtained through simple computation (e. g. growth rate as percentage change in confirmed cases). Other analyses used simple modeling approaches, e. g. fitting an exponential growth model to the time series and extracting the exponential growth rate or doubling time from the estimated parameters.
(2) Epidemiological parameters were mostly estimated from confirmed cases or deaths. Some approaches fitted a compartmental transmission model to the raw epidemiological outcome. For this, the parameter of interest was either allowed to vary over time, or the model was fitted independently on different time periods. Other approaches employed a statistical method to directly estimate reproduction numbers from the observed outcome. Here, the method by Cori et al. [27] as implemented in the popular software package "EpiEstim" [28] for estimation of the instantaneous effective reproduction number was used in a large number of analyses. However, we found that statistical methods were not always applied correctly, which could have led to bias in the inferred transmission dynamics (see online Appendix A). Sometimes, authors also used methods to estimate reproduction numbers from contact matrices [29] (derived from surveys on personal contacts) or from transmission chains [30, 31] (derived from contact tracing data).
(3) Summary statistics were typically obtained through simple computation.
(4) Change points in the outcome were obtained by fitting a compartmental transmission model with special parameters representing points in time when the transmission rate changes [26]. Other analyses used special change point detection algorithms [25].

Data source

The majority of authors directly accessed surveillance data from national health authorities or other governmental bodies. In the case of individual-level data, which may be subject to privacy regulations, authors were often themselves affiliated to the relevant health authority. To obtain population-level data, a considerable share of analyses also used publicly available data from cross-country selections, e. g. the European Centre for Disease Prevention and Control (ECDC) [32], the Johns Hopkins University (JHU) [33], or Worldometer [34], which offer aggregated surveillance data internationally from various sources for the pandemic. Mobile phone tracking data were usually provided by corporate organizations such as Google [35], Apple [36], or Baidu [37]. A few analyses were also based on data collected by the authors, e. g. survey data on behavioral outcomes, seroprevalence studies, or data collected at a local facility such as a hospital.

Data availability

Data for the raw outcome was usually publicly available, in particular for epidemiological population-level outcomes such as cases and deaths because such data could oftentimes be accessed via the source that is documented in the manuscript. In several cases, the data was made publicly available by the study authors, e. g. by depositing the analyzed data in a public repository. For a small, yet considerable number of analyses, data was not accessible as the data was neither made publicly available nor the source of the data could be identified. Of note, data on epidemiological individual-level data was typically not available due to privacy concerns. Furthermore, corporate mobility data was widely available in the past, but access has recently been restricted by many providers.

Intervention

The analyses vary in the types of exposures and non-pharmaceutical interventions. A systematic classification is shown in Table 3.

Terminology for non-pharmaceutical interventions

Varying terminology was used by the literature to refer to non-pharmaceutical interventions (Table 3, D.3.1 and D.3.2). This is reflected in our search string, where we used a large set of terms in order to capture a broad range of relevant studies. While terminology sometimes reflected the specific types of non-pharmaceutical interventions that were analyzed, differences in terminology may also be the result of different research backgrounds of the study authors.

Exposure types and types of single interventions

A considerable number of analyses examined one single [5, 38] or multiple interventions separately [2, 7]. Among these analyses, school closures and stay-at-home orders were examined most frequently, which may be due to these interventions being particularly controversial in the public

discourse [39, 40]. The majority of analyses, however, did not examine multiple interventions separately but rather analyzed the a combination of multiple interventions jointly [41–43], which is often the case when multiple interventions were implemented on the same day and when thus the separate associations of the outcome with interventions could not be disentangled. A considerable number of analyses were even less specific by only analyzing whether interventions were altogether effective but without attributing changes in the outcome to specific interventions [44, 45]. Other ways to assess the effectiveness of interventions were: examining the start time of interventions [26, 46], e. g. to compare different delays with which governments responded to the pandemic [46]; dividing the public health response into different periods [47, 48]; dividing interventions into different categories [49, 50]; or summarizing the stringency of interventions to a numerical index at a specific time point [51, 52].

Of note, analyses that examined a combination of interventions often referred to this combination as "lockdown". In the underlying analyses, such lockdowns typically included multiple interventions implemented on the same day [42, 53]. However, the specific interventions included in lockdowns varied considerably between populations. We therefore considered "lockdown" as an umbrella term for different combinations of interventions rather than as a specific type of intervention. Furthermore, some studies did not only assess the relationship between mobility and non-pharmaceutical interventions, but also between changes in mobility and population-level epidemiological outcomes. In these analyses, human mobility was typically defined as a continuous exposure. We extracted information on such complementary analyses of mobility as an addendum to the main review (see online Appendix E).

Coding of interventions

When multiple populations were jointly analyzed, coding of interventions may have been necessary in order to reconcile differences in the definitions of interventions between populations. For instance, the term "school closures" could refer to the closure of primary or secondary schools or universities. Differences across populations are thus reconciled during coding by deciding upon the type of intervention and providing a common name and definition that is then applied to all populations. As a result, such coding can be subjective and thus needs to be carefully documented and evaluated (see online Appendix A). Coding of interventions was necessary in around a quarter of analyses.

Source of intervention data and availability of data on exposure

If coding of interventions was *not* necessary, authors often obtained intervention data (i. e. the date of interventions) from a government or news website. Unfortunately however, the data source was often not provided by the authors and could thus not be evaluated. If coding of interventions was necessary, then study authors either coded the data themselves, i. e. collected the data from government or news websites and systematically categorized them or, more frequently, used externally coded data. The most popular choices for externally coded data were the Oxford Government Response Tracker [1] and, for the United States, the New York Times [54].

Methodological approach

A variety of methodological approaches were used to assess the effectiveness of interventions. The methodological approaches extracted here describe the actual stage of estimating the associations of health-related outcomes with interventions. A systematic classification of the methodological approaches is shown in Table 4. An additional analysis of the average citation count per category is presented in online Appendix C.

Empirical approach

We distinguished three general empirical approaches for assessing the effectiveness of interventions, namely **(D)** descriptive, **(P)** parametric, and **(C)** counterfactual approaches.

- **(D)** Descriptive approaches were used by the majority of analyses: These approaches provided descriptive summaries of the outcome over time or between populations, and related variation in these summaries to the presence or absence of different interventions. For example, some analyses compared changes in the growth rate of observed cases before and after interventions were implemented [55, 56]. Of note, descriptive approaches could involve modeling as part of an intermediate step, where a latent outcome was computed from the raw outcome (see Computed outcome), while, afterward, a descriptive approach was used to the link the latent outcome to interventions. For example, some analyses used a single-population compartmental trans-

Table 3 Systematic classification and frequency of the interventions (D.3)

D.3.1: Terminology for interventions[†‡]	Frequency
Not applicable (only specific term for intervention type)	22 (9%)
Measures	135 (54%)
Interventions	65 (26%)
Policies	16 (6%)
Other	14 (6%)
D.3.2: Terminology for the specific type of non-pharmaceutical interventions[†‡]	Frequency
Not applicable (only general term for interventions)	3 (1%)
Non-pharmaceutical	49 (16%)
Control	48 (16%)
Social distancing	45 (15%)
Other	159 (52%)
D.3.3: Exposure types	
One single intervention	43 (15%)
Multiple separate interventions	31 (11%)
One combination of interventions	84 (29%)
Multiple combinations of interventions	20 (7%)
All interventions together	70 (25%)
Other	37 (13%)
D.3.4: Types of single interventions	
Not applicable (no single interventions analyzed)	211 (74%)
One or multiple single interventions analyzed (as defined in D.3.4 of the Documentation manual)[‡]	74 (26%)
Stay-at-home order	44 (59%)
Other	27 (36%)
School closure	25 (34%)
Workplace closure	20 (27%)
International travel restrictions	17 (23%)
Declaration of a state of emergency	13 (18%)
Bans of large gatherings	13 (18%)
Venue closure	12 (16%)
Bans of small gatherings	10 (14%)
D.3.5: Coding of interventions	
D.3.6: Source of intervention data	
Not applicable (no specific interventions analyzed)	74 (26%)
Not necessary (no joint analysis of interventions across multiple populations)	137 (48%)
Could not be evaluated	98 (72%)
Government or news websites	30 (22%)
Other	9 (7%)
Necessary (joint analysis of interventions across multiple populations)	74 (26%)
Could not be evaluated	9 (12%)
Coding done by authors	20 (27%)
Use of externally coded data	45 (61%)
D.3.7: Availability of data on exposure	
Not applicable (no specific interventions analyzed)	73 (26%)
Raw data documented in the manuscript	136 (48%)
Access to externally coded data via source	32 (11%)
Coded datamade available by the authors	34 (12%)
Coded data not available	10 (4%)

[†] Results for this subdimension are reported at the study-level, and not the level of analysis (i. e. one study can contain multiple analyses). If a study uses more than one term predominantly, then both are counted and added to the total count.

[‡] Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100 %

Table 4 Systematic classification and frequency of the methodological approach (D.4)

D.4.1: Empirical approach				Total freq.
D: Descriptive				151 (53%)
P: Parametric				94 (33%)
C: Counterfactual				40 (14%)
D.4.2: Use of exposure variation	**(D)**	**(P)**	**(C)**	
Only variation over time for a single population	78	23	24	125 (44%)
Only variation over time for multiple populations	63	22	10	95 (33%)
Only variation between populations	4	14	0	18 (6%)
Both variation over time and between populations	6	35	6	47 (16%)
D.4.3: Method				
Description of change over time	136	—	—	136 (48%)
Description of time course				49 (36%)
Comparison of time periods				87 (64%)
Comparison of populations	8	—	—	8 (3%)
Comparison of change points with intervention dates	7	—	—	7 (2%)
Non-mechanistic model	—	61	17	78 (27%)
Generalized linear model				51 (65%)
Exponential growth model				11 (14%)
Other				16 (21%)
Mechanistic model	—	30	13	43 (15%)
Compartmental single-population transmission modl				29 (67%)
Compartmental meta-population transmission model				4 (9%)
Semi-mechanistic Bayesian transmission model				5 (12%)
Other				5 (12%)
Synthetic controls	—	—	6	6 (2%)
Other	0	3	4	7 (2%)
D.4.4: Code availability				
None (not available)	121	66	33	220 (77%)
Publicly available	30	28	7	65 (23%)

Empirical approach: **(D)** descriptive, **(P)** parametric, and **(C)** counterfactual

mission model to estimate the time-varying reproduction number and then compared the reproduction number before and after interventions were implemented [57–59].

- **(P)** Parametric approaches were used by a third of analysis: These approaches formulated an explicit link between intervention and outcome, where the association was quantified via a parameter in a model. Most frequently these were regression-like links between interventions and the reproduction number [2, 8].
- **(C)** Counterfactual approaches were least frequently used: These approaches assessed the effectiveness of interventions by comparing the observed outcome with a counterfactual outcome based on an explicit scenario in which the interventions were not implemented. For example, the observed number of cases was compared with the number of cases that would have been observed if the exponential growth in cases had continued as before the implementation of interventions [60, 61].

Use of exposure variation

Effectiveness of interventions were assessed by exploiting variation in the exposure to the intervention over time, between populations, or both. Assessments exploiting exposure variation over time contrasted the outcome in time periods when specific measures were in place with the outcome in time periods when they were not in place. In contrast, assessments exploiting exposure variation between populations were based on a comparison of the outcome between populations that were subject to specific measures with populations that were not. Only a small share of analyses exploited variation between populations or both between populations and over time.

Method

We grouped the different methods used into (1) description of change over time, (2) comparison of populations, (3) comparison of change points with intervention dates, (4) non-mechanistic model, (5) mechanistic model, and (6) synthetic controls. We review these in the following.

(1) *Description of change over time*

The large majority of analyses following a descriptive approach examined the change of the outcome over time to assess the effectiveness of interventions. In some of these analyses, the focus was on the course of the outcome over time, typically by attributing the observed change (e. g. a reduction in new cases over time) to the analyzed interventions. For example, the outcome was assessed at regular or irregular intervals, which were not necessarily aligned with the implementation dates of interventions [44, 45, 62]. The majority of analyses, however, followed the logic of an interrupted time series analysis, i. e. the outcome was explicitly compared between time periods before and after interventions [63–66].

(2) *Comparison of populations*

A few descriptive analyses compared outcomes via summary statistics only between populations (i. e. without considering variation over time) to assess the effectiveness of interventions. In such analyses, the outcomes were compared between populations that were stratified by different exposure to interventions (e. g. populations that implemented a certain intervention and populations that did not) [67–69].

(3) *Comparison of change points with intervention dates*

Some descriptive analyses checked whether the dates of estimated change points in outcomes and the implementation dates of interventions coincide [10, 25, 70]. If both dates were more or less in agreement, this was taken as evidence confirming the effectiveness of the intervention. However, change point detection methods could also yield change points prior to the implementation of interventions, which was sometimes interpreted as a sign of additional factors influencing the outcome (e. g. proactive social distancing) [25].

(4) *Non-mechanistic model*

Non-mechanistic models are statistical models that typically make *no* explicit assumptions about the mechanisms that drive infection dynamics. Such models were used in both parametric and counterfactual approaches by a quarter of analyses.

In parametric approaches, non-mechanistic models—almost always (generalized) linear regression models—were used to model a direct link between interventions and outcome. Typically, dummy variables were used to indicate when (variation over time) [9, 71, 72] or where (variation between populations) [73–75] interventions were implemented. Analyses exploiting both variation over time and between populations typically used panel regression methods [5, 76, 77].

In counterfactual approaches, the non-mechanistic models used were mostly exponential growth models, and sometimes time series models (e. g. AR(I)MA or exponential smoothing) [41, 47, 61]. These models were fitted using data prior to when an intervention was implemented and then extrapolated the outcome afterwards.

(5) *Mechanistic model*

Mechanistic models have a structure that makes, to some extent, explicit assumptions about the mechanisms that drive infection dynamics. They were used in both parametric and counterfactual approaches by slightly more than ten percent of analyses.

In parametric approaches, the association of an outcome with an intervention was represented via a parameter that was functionally linked to the disease dynamics (i. e. via a latent variable) of the model. This was typically achieved by parameterizing the transmission rate or reproduction number as a function of binary variables, indicating whether interventions were implemented or not [2, 78–80]. Others linked interventions to the contact rate, the transmission probability upon contact, or to entries in the contact matrix [81–83]. A few modeling approaches also represented the intervention via an explicit structure or dynamic in the model, e. g. a compartment for quarantined individuals with a quarantine rate [50, 84] or an exponential decay of the susceptible population [49, 50, 85].

The most popular mechanistic models used in parametric approaches were compartmental transmission models. These models were fitted to the time series of cases, hospitalizations, recovered cases, deaths, or several simultaneously. With the exception of one meta-population model [86], all compartmental models used in analyses following a parametric approach were single-population models. If multiple populations were analyzed, each population was modeled separately. A few parametric analyses also used a semi-mechanistic Bayesian transmission model with a time-discrete renewal process, similar to the one in an early influential paper by Flaxman et al. [8]. These analyses fitted a Bayesian hierarchical model with stochastic elements for disease transmission and

ascertainment on observed time series for cases, deaths, or both [2, 8, 87]. The model was usually fitted to data from several populations, modeling separately the time course in each population but estimating the parameters for the associations of outcome with interventions jointly across populations. Rarely, analyses used highly complex models such as individual-based transmission models simulating the behavior of individual agents, or phylodynamic models inferring both virus phylogenies and transmission dynamics from genome sequence data.

In counterfactual approaches, mechanistic models were, similar to non-mechanistic models, calibrated to data before the implementation of an intervention and then projected the outcome for the time after the intervention, while keeping the model parameters fixed [88–90]. Thus, no relationship between intervention and outcome is explicitly modeled. Regularly, these analyses used meta-population or individual-based models that incorporated migration dynamics through mobility data and a network between individuals or populations [90–92].

(6) *Synthetic controls*

Some counterfactual approaches used synthetic control methods. Here, a counterfactual scenario was constructed by computing the counterfactual outcome as a weighted combination of observations from a pool of "control" populations in which the intervention was not implemented [46, 93, 94]. Weights were fitted so as to give more importance to control populations similar to the intervention population. In these analyses, the course of the outcome before intervention was often used as the primary measure of similarity [6, 93, 94]. Sometimes, further factors such as geographic proximity or population characteristics were also considered [93, 95].

Code availability

For around one in four analyses, a link to a publicly accessible repository containing the computer code implemented for a specific analysis was provided. Overall, the code availability was comparably higher for parametric approaches, where one in three analyses provided a link.

Effectiveness assessment

The analyses in our review sample varied in their form of effectiveness assessment, i. e. how the association of outcomes with interventions were quantified, whether they were interpreted causally, whether uncertainty was reported, and whether sensitivity analyses or subgroup assessments were conducted. A systematic classification of the effectiveness assessment is shown in Table 5.

Reporting of effectiveness, measure of effectiveness, and reporting of uncertainty

Around one in five analyses qualitatively described the change in the outcome over time following the implementation of interventions. More frequently the outcome values before an intervention were compared with the outcome values after an intervention. Around half of the analyses reported a quantitative change in outcome values, e. g. by computing the difference in the outcome values before and after an intervention, or estimating the difference via a parameter in a statistical model. The effectiveness was oftentimes measured in terms of a change in the reproduction number, in confirmed cases, or in mobility, but many other measures of effectiveness were also common. Uncertainty was reported in around one half of the analyses, e. g. via standard error, confidence intervals, and credible intervals.

Interpretation of results

Some analyses, in particular those describing the change of the outcome over time, interpreted their results only as associative [63, 77], i. e. a statistical or temporal association between interventions and the measure of effectiveness was noted without a causal implication. In the majority of analyses, however, a causal conclusion was implicitly drawn from the results through the use of causal language (e. g. it was concluded that "interventions reduced transmission") [8, 44]. Only rarely, and mostly in analyses using non-mechanistic econometric models [96, 97] or synthetic controls, results were explicitly described as estimates of the causal effects of interventions.

Sensitivity analyses

We checked all works for sensitivity analyses that were specifically conducted to examine the robustness of the reported effectiveness. Many studies conducted sensitivity analyses only related to the predicted outcome or model fit, but not to the effectiveness of interventions. Overall, the vast majority of analyses did not conduct sensitivity analyses with regard to the effectiveness.

Of those that did, sensitivity analyses focused on model extensions or adjustments in which the model specification was varied, e. g. by changing the structure of a transmission model or by adjusting the estimated effects of interventions for additional variables in a regression

Table 5 Systematic classification and frequency of different effectiveness assessments (D.5)

D.5.1: Reporting of effectiveness				Total freq.
QS: Qualitative statement				53 (19%)
CO: Comparison of outcome values				73 (26%)
QC: Quantification of change in outcome values				159 (56%)
D.5.2: Measure of effectiveness‡	**(QS)**	**(CO)**	**(QC)**	
Change in reproduction number	22	44	29	95 (33%)
Change in confirmed cases	16	15	38	69 (24%)
Change in mobility	9	6	28	43 (15%)
Other	18	29	100	147 (52%)
D.5.3: Interpretation of results				
Associative				111 (39%)
Implicitly causal				160 (56%)
Explicitly causal				14 (5%)
D.5.4: Reporting of uncertainty				
Not applicable				52 (18%)
Yes				154 (54%)
No				79 (28%)
D.5.5: Sensitivity analysis (including computed outcomes)				
None (no sensitivity analyses w.r.t effect)				217 (76%)
One ore more sensitivity analyses‡				68 (24%)
Model specification varied				36 (53%)
Epidemiological parameters varied				29 (43%)
Different or modified outcome used				17 (25%)
Same analysis with (sub)population excluded				16 (24%)
Different coding of interventions used				10 (15%)
Start or end date of study period varied				4 (6%)
D.5.6: Subgroup assessment				
None (no subgroups)				250 (88%)
One or more subgroups‡				35 (12%)
Based on socioeconomic indicators				23 (66%)
Based on epidemiological indicators				16 (46%)
Based on public health response				9 (26%)
Based on geographic areas				6 (17%)

Reporting of effectiveness: **(QS)** qualitative statement, **(CO)** comparison of outcome values, and **(QC)** quantification of change in outcome values

‡ Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100%

model. Others analyzed sensitivity with respect to variations in epidemiological parameters, e. g. by assuming a different basic reproduction number, generation or serial interval, infectious period, or reporting delay distribution. Only few analyses tested sensitivity with regard to data: i. e. using different or modified outcomes [72, 98]; using a different coding of interventions [3, 7]; or repeating the same analysis but excluding (sub)populations [2].

Subgroup assessment

The effectiveness of interventions were rarely assessed within subgroups of the population. Two thirds of such assessments were within subgroups created based on socioeconomic indicators, e. g. by age and gender [44] or by regions with different income levels [6] Less frequent were subgroups based on epidemiological indicators [5], the public health response [99], or geographic areas [100].

Discussion

Our systematic review covers over 240 studies published between January 2020 and January 2021. Insights from this review can inform different types of future studies: (1) studies using data from the same period that extend our knowledge on aspects that have so far been rarely investigated; (2) studies using data from subsequent periods that generate new insights or corroborate existing ones; and (3) studies using data from a future pandemic caused by another virus. Although the preconditions to conduct these studies differ, they share the goals and challenges of the studies in our review sample. Accordingly, the results from our systematic review allow us to discuss implications for future work and make recommendations for improving methodologies and comparability across studies.

Implications for future work

During the COVID-19 pandemic, both surveillance data on confirmed cases, hospitalizations, or deaths [32, 33], and mobility data from mobile phones [35, 36] have become publicly available at scale. This has enabled a large number of studies assessing the associations of population-level epidemiological outcomes and human mobility with non-pharmaceutical interventions (Table 2, D.2.1). However, considerable potential remains in the exploration of outcomes and analyses that have so far been rarely employed.

First, the population-level data used by the majority of studies in our review sample can be subject to systematic differences in ascertainment between populations and over time. For example, epidemiological analyses have discussed the influence of testing procedures and intensity on the number of confirmed cases [2, 14, 42]. Due to limited availability of metadata from health authorities, it is oftentimes difficult to account for such factors. In this context, smaller-scale surveys with precisely defined outcomes and controlled sampling schemes (e. g. representative community sampling [101]) could provide a complementary source of data for future studies. Similarly, as has been demonstrated by studies in our review sample, the use of individual-level data could allow for more detailed analyses, e. g. by relating non-pharmaceutical interventions to changes in the serial interval using symptom onset data [44], to transmission chains using contact tracing data [90], or to virus migration rates using genome sequence data [23]. We hope to see more such analyses as more individual-level data becomes available.

Second, there can be great merit in analyses advancing our understanding of the mechanisms by which non-pharmaceutical interventions work. For example, interventions may influence behaviour and transmission through factors not captured by previous studies using mobility data from mobile phones, and, moreover, the relationship between population behavior and disease transmission may change over time [102, 103]. Additional insights can be gained from analyses using behavioral data from other sources, e. g. surveys evaluating compliance with mask mandates [65] or the number of daily contacts [104]. Moreover, we see value in analyzing interventions, behavior and epidemiological outcomes jointly, i. e. in the form of a mediation analysis [96, 105], allowing to differentiate the direct and indirect effect of non-pharmaceutical interventions.

Third, only one in ten analyses in our review sample examined variation in the effectiveness of interventions across subgroups or populations (Table 4, D.4.2). Estimating and explaining such variation could help understand the conditions under which interventions are more or less effective for a specific subgroup, potentially allowing policy makers to tailor interventions to a specific subgroup or setting. Our review points out two approaches to analyze such variation: (1) comparing the effectiveness of interventions between subgroups of the same population (e. g. between the young and elderly population)[44]; and (2) comparing the effectiveness of interventions between different populations and relating differences to population-specific characteristics (e. g. population density) [106].

Last, while many analyses in our review sample used terminology related to a causal interpretation of the provided evidence (Table 5, D.5.3), it is difficult in general to estimate causal effects based on population-level observational data and to rule out unobserved confounding, e. g. from voluntary behavioral changes [107–110]. Hence, the evidence from the studies in our review sample should generally be interpreted as associative rather than causal estimates. A causal interpretation may be justified when additional criteria are met (e. g. the "no unmeasured confounding" assumption [111] or the Bradford Hill criteria [112] and extensions thereof [113–115]), but until now a rigorous discussion of such assumptions when estimating the effects of non-pharmaceutical interventions is lacking. Apart from that, we caution against emphasizing results from single studies. Rather, we recommend evaluating evidence from multiple observational studies jointly and, more importantly, in combination with other types of evidence. For example, evidence on the infectiousness of school children and parental strategies to fill the care gap can produce independent predictions about the effectiveness of school closures. Similarly, laboratory evidence regarding the effectiveness of masks together with evidence on compliance with masks can produce independent predictions of the effectiveness of mask mandates.

Recommendations for improving methodologies

Variation in the exposure to interventions (i. e. when, where and which interventions were implemented) is required in order to empirically assess their effectiveness. However, changes in the outcome over time may falsely be attributed to non-pharmaceutical interventions if they are subject to confounding by concurring time trends. We thus recommend to also exploit exposure variation between populations, i. e. with respect to the timing and the types of single interventions that were implemented. This was done by only one in five analyses in our review sample (Table 4, D.4.2), although we found that on average these studies had more citations than other studies (see online Appendix C). Given that the types and timing of interventions varied considerably between populations, a valuable source of variation remains largely untapped by most analyses.

Evidenced-based decision making requires empirical estimates for the effects of single non-pharmaceutical interventions (e. g. school closures or stay-at-home orders). However, the majority of analyses assessed the effectiveness of population-specific combinations of interventions such as lockdowns (Table 3, D.3.3). The underlying analyses typically studied only a single population (or multiple populations separately) where multiple interventions were implemented on the same day, and, as a result, the separate associations of outcomes with interventions cannot be disentangled. For future work, we recommend more effort to conduct analyses across multiple populations, so that the separate associations of outcomes with single interventions can be dissected.

Recommendations for improving comparability across studies

During a pandemic, public health policy has a strong focus on the number of confirmed cases, hospitalizations, and deaths, making them obvious outcomes to evaluate the effectiveness of interventions. However, non-pharmaceutical interventions act only indirectly and with a certain delay on these observable outcomes. Typically, non-pharmaceutical interventions should influence the behaviour of the population, which should reduce transmission (e. g. by limiting the contact rate), which in turn should affect the number of new infections and, subsequently, observed outcomes like confirmed cases, hospitalizations, or deaths. The question of how to assess the effectiveness of interventions along this path has been answered differently by the studies in our review sample. We identified four main types of analyses; see (1)–(4) in Box 1. In the following, we discuss the different types with regard to their ability of enabling a comparison of results between studies.

Box 1. Different types of analyses to assess the effects of non-pharmaceutical interventions

(1) Observed outcome directly linked to interventions	A raw, observed outcome is analyzed directly by evaluating differences (1) over time with an interrupted time-series analysis comparing the outcome before vs. after an intervention, (2) between populations with a cross-sectional analysis comparing populations exposed vs. not exposed to an intervention, or (3) both over time and between populations with a panel data analysis. Mechanistic modeling is typically not involved in this type of analysis, with one exception, namely counterfactual approaches using a transmission model to project the observed outcome after intervention.
(2) Computed, unobserved outcome linked to interventions	In contrast to type (1), the intervention effect is measured in terms of an unobserved outcome. This is computed from the raw outcome and then analyzed in a similar manner as in (1). Mechanistic modeling can be involved in computing the unobserved outcome, for example by using a model to estimate the reproduction number or transmission rate from the number of new cases.
(3) Observed outcome linked to interventions via unobserved outcome in mechanistic model	Observed outcomes are used to fit a mechanistic model (e. g. compartmental transmission model) that includes a latent variable representing an unobserved outcome (e. g. the reproduction number), which in turn is parameterized as a function of interventions. For instance, a regression-like link is used within the mechanistic model to estimate the effect of interventions on the transmission rate as a latent variable.
(4) Change points in outcome related to exposure	Change points are estimated in the time series of an observed or unobserved outcome. The estimated change points are then related to the implementation dates of interventions. If the estimated change points agree well with the actual implementation dates of interventions, this is interpreted as evidence for the effectiveness of interventions.

Analyses of type (1) can avoid mechanistic modeling by directly analyzing an observed outcome such as cases or deaths. Here, a central challenge is to take into account the uncertain delay between the implementation of non-pharmaceutical interventions and their effects on the observable outcome. The fact that infections and subsequent outcomes such as confirmed cases follow exponential dynamics during an epidemic wave makes it difficult to compare estimates measured by observable outcomes across different epidemic phases. In contrast to that, analyses following type (2) or (3) employ mechanistic modeling, allowing to link latent, unobservable outcomes (e. g. the transmission rate or the reproduction number) to interventions. Since these latent outcomes can be inferred from different observed outcomes like cases or deaths, it becomes possible to compare analyses that use different raw data. The difference between type (2) and (3) is that for (2) the estimation of the latent outcome is separated from the effectiveness assessment. Such separation reduces model complexity, however, often at the expense of incomplete uncertainty assessments (if uncertainty regarding the computed outcome is left out). Finally, analyses following type (4) take a very different approach that shares few assumptions with the other approaches: A comparison of change points can verify the presence of an association, yet without quantifying its size. As a result, such findings are best complemented with an analysis of type (1), (2), or (3). Notably, we found that studies which received many citations often used analyses of type (2) or (3) (see online Appendix C).

While variation in methodologies can complicate the comparison of studies, it may help to identify the influence of certain methodological choices on the results. Here, the public availability of data for outcomes and interventions holds potential for sensitivity analyses within studies as well as comparisons between studies. Specifically, the same analysis could be repeated with different sets of publicly available data as part of the same study. This way, sensitivity of the findings with respect to the choice of outcome and intervention data could be assessed within studies, reducing the risk of bias from specific outcome data (e. g. incomplete case ascertainment due to limited testing capacity etc.) or the specific coding of interventions. For example, the number of new cases, deaths, or both could be used as the raw outcome in mechanistic models with a comparable latent outcome [2]. However, other aspects, in particular the specific setting and methodologies used, are presumably more difficult to vary as part of a sensitivity analysis, and may therefore need to be compared between different studies. Important for such comparisons is giving access to the preprocessed data even if the raw data was retrieved from public sources (see online Appendix A).

Limitations of our review

Our systematic review has limitations. First, it covers studies published between January 1, 2020 and January 12, 2021. While comprising a large review sample of more than 240 publications, it is therefore limited to the first year of the COVID-19 pandemic. A future review may examine to what extent methodologies have changed over time or what novelties were introduced when analyzing data from later waves, although we have referred to some recent contributions in our discussion. Second, although our review process aimed to ensure a representative sample of studies in the field, certain biases cannot be ruled out. Specifically, our search queries focused on general terms describing non-pharmaceutical interventions (see Methods), so that studies using only terminology related to specific interventions may not have been found. Third, our data extraction form comprises items that were widely applicable over a diverse set of studies. While providing a consistent framework to compare different methodologies, this naturally limits in-depth analyses of specific methodologies, which could complement our review.

Comparison with related work

Our review provides the first large, systematic categorization of existing methodologies to assess the effectiveness of non-pharmaceutical interventions. An early review on the subject was written by Perra [16], which is however not systematic and has a different objective than our methodology review (i. e. the majority of studies in the review sample cover other aspects than the effectiveness of non-pharmaceutical interventions). A more recent methodology review by Garin et al. [116] focuses on epidemic models during the COVID-19 pandemic. It thus complements our review by considering a subset of the methodologies in our review sample and studying them beyond their use in assessing the effectiveness of non-pharmaceutical interventions. Finally, besides methodology reviews, several meta-analyses have attempted to summarize the effectiveness of non-pharmaceutical interventions [11–15].

Conclusions

Our review of more than 240 studies on the effectiveness of non-pharmaceutical interventions revealed substantial variation in methodologies. Until specific best practices emerge, further heterogeneity in studies is inevitable and can also be beneficial, e. g. for assessing robustness of the results with respect to method and input data. Nevertheless, some

standardization is required in order to synthesize evidence on the effectiveness of non-pharmaceutical interventions from multiple studies. So far, a lack of common standards and substantial variation in the methodologies used have created a challenge for meta-analyses to summarize and compare the reported evidence from existing studies [11–15]. Here, our methodology review can serve as a basis for subsequent meta-analyses to factor in the variety of existing methodologies when pooling and comparing the reported evidence. Moreover, the systematic categorization of methodologies developed in this review may also serve as a basis for designing a risk of bias assessment tool specific to studies on the effectiveness of non-pharmaceutical interventions. Most importantly though, our recommendations for the design of future studies aim to extend the scope of existing analyses and reduce methodological barriers to comparability across studies.

During the COVID-19 pandemic, a tremendous amount of publicly available epidemiological data has been generated. The ease of access to this data allowed many researchers to contribute work, using a variety of methodologies to assess the effectiveness of non-pharmaceutical interventions on health-related outcomes. With researchers from diverse fields contributing, there is a unique opportunity to benefit from the various inputs in developing a methodological foundation for timely and robust assessments during future pandemics. This will however require a thorough examination of the present methodologies in order to share lessons learnt and develop best practices. Our systematic review can be viewed as a first such attempt.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10654-022-00908-y.

Author contributions NB and AL contributed to conceptualization, literature search, study selection, data extraction, data synthesis, and manuscript writing, review & editing. DO contributed to literature search and manuscript review & and editing. TS contributed to manuscript review & editing. SF contributed to manuscript writing, review & editing. WV contributed to conceptualization, study selection, data extraction, data synthesis, manuscript writing, review & editing.

Funding Open access funding provided by Swiss Federal Institute of Technology Zurich. NB and SF acknowledge funding from the Swiss National Science Foundation (SNSF) as part of the Eccellenza grant 186932 on "Data-driven health management". The funding bodies had no control over design, conduct, data, analysis, review, reporting, or interpretation of the research conducted.

Data availibility The full data extracted from the studies in this review is available in machine-readable format at https://github.com/adrian-lison/methodologies-npi-effects.

Declarations

Competing interests All authors declare no competing interests.

Ethics approval Ethics approval was not required for this study.

References

1. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). Nat Hum Behav. 2021;5(4):529–38.
2. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. Science. 2021;371(6531):eabd9338.
3. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Hum Behav. 2020;4(12):1303–12.
4. Banholzer N, van Weenen E, Lison A, Cenedese A, Seeliger A, Kratzwald B, et al. Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. PLoS ONE. 2021;16(6): e0252827.
5. Auger KA, Shah SS, Richardson T, Hartley D, Hall M, Warniment A, et al. Association between statewide school closure and COVID-19 incidence and mortality in the US. JAMA. 2020;324(9):859–70.
6. Bennett M. All things equal? Heterogeneity in policy effectiveness against COVID-19 spread in Chile. World Dev. 2021;137: 105208.
7. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. Health Aff. 2020;39(7):1237–46.
8. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584(7820):257–61.
9. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature. 2020;584(7820):262–7.
10. Lemaitre JC, Perez-Saez J, Azman AS, Rinaldo A, Fellay J. Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland. Swiss Med Wkly. 2020;150: w20295.
11. Mendez-Brito A, Bcheraoui CE, Pozo-Martin F. Systematic review of empirical studies comparing the effectiveness of non-pharmaceutical interventions against COVID-19. J Infect. 2021;83(3):281–93.
12. Poeschl J, Larsen RB. How do non-pharmaceutical interventions affect the spread of COVID-19? A literature review. Danmarks Nationalbank; 2021. 4.

13. Rizvi RF, Craig KJT, Hekmat R, Reyes F, South B, Rosario B, et al. Effectiveness of non-pharmaceutical interventions related to social distancing on respiratory viral infectious disease outcomes: a rapid evidence-based review and meta-analysis. SAGE Open Med. 2021;9.
14. Iezadi S, Gholipour K, Azami-Aghdash S, Ghiasi A, Rezapour A, Pourasghari H, et al. Effectiveness of non-pharmaceutical public health interventions against COVID-19: a systematic review and meta-analysis. PLoS ONE. 2021;16(11): e0260371.
15. Talic S, Shah S, Wild H, Gasevic D, Maharaj A, Ademi Z, et al. Effectiveness of public health measures in reducing the incidence of COVID-19, SARS-CoV-2 transmission, and Covid-19 mortality: systematic review and meta-analysis. BMJ. 2021;375: e068302.
16. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. Phys Rep. 2021;913:1–52.
17. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane handbook for systematic reviews of interventions. Cochrane; 2021.
18. Banholzer N, Lison A, Özcelik D, Feuerriegel S, Vach W. A comparison of studies estimating the effectiveness of non-pharmaceutical interventions: a systematic review protocol. PROSPERO; 2021.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.
20. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007;18(6):805–35.
21. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLoS Med. 2015;12(10): e1001885.
22. Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a reporting guideline for population health and policy interventions. BMJ. 2018;361: k1079.
23. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science. 2020;369(6508):1255–60.
24. Moreno GK, Braun KM, Riemersma KK, Martin MA, Halfmann PJ, Crooks CM, et al. Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. Nat Commun. 2020;11:5558.
25. Wieland T. A phenomenological approach to assessing the effectiveness of COVID-19 related nonpharmaceutical interventions in Germany. Saf Sci. 2020;131: 104924.
26. Karnakov P, Arampatzis G, Kii I, Wermelinger F, Wlchli D, Papadimitriou C, et al. Data-driven inference of the reproduction number for COVID-19 before and after interventions for 51 European Countries. Swiss Med Wkly. 2020;150: w20313.
27. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178(9):1505–12.
28. Cori A. EpiEstim: estimate time varying reproduction numbers from epidemic curves; 2021.
29. Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. J Math Biol. 1990;28(4):365–82.
30. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160(6):509–16.
31. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–9.
32. European Centre for Disease Prevention and Control. COVID-19 Datasets; 2022. https://www.ecdc.europa.eu/en/covid-19/data.
33. Johns Hopkins University & Medicine. Coronavirus Resource Center; 2022. https://coronavirus.jhu.edu/.
34. Worldometer. Coronavirus Statistics; 2022. https://www.worldometers.info/coronavirus/.
35. Google. COVID-19 Community Mobility Reports; 2022. https://www.google.com/covid19/mobility/.
36. Apple. COVID-19 Mobility Trends Reports; 2022. https://covid19.apple.com/mobility.
37. China Data Lab. Baidu Mobility Data; 2021. https://doi.org/10.7910/DVN/FAEZIO.
38. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;20(5):553–8.
39. Couzin-Frankel J, Vogel G. School openings across globe suggest ways to keep coronavirus at bay, despite outbreaks. Science; 2021.
40. Berry CR, Fowler A, Glazer T, Handel-Meyer S, MacMillen A. Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic. Proc Natl Acad Sci. 2021;118(15): e2019706118.
41. Bönisch S, Wegscheider K, Krause L, Sehner S, Wiegel S, Zapf A, et al. Effects of coronavirus disease (COVID-19) related contact restrictions in Germany, March to May 2020, on the mobility and relation to infection patterns. Front Public Health. 2020;8: 568287.
42. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020;368(6490):493–7.
43. Salvatore M, Basu D, Ray D, Kleinsasser M, Purkayastha S, Bhattacharyya R, et al. Comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: national trends masking state-level variations. BMJ Open. 2020;10(12): e041778.
44. Ali ST, Wang L, Lau EHY, Xu XK, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369(6507):1106–9.
45. Price DJ, Shearer FM, Meehan MT, McBryde E, Moss R, Golding N, et al. Early analysis of the Australian COVID-19 epidemic. eLife. 2020;9: e58785.
46. Huber M, Langen H. Timing matters: the impact of response measures on COVID-19-related hospitalization and death rates in Germany and Switzerland. Swiss J Econ Stat. 2020;156(1):1–19.
47. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. Lancet Digit Health. 2020;2(12):e638–49.
48. Jefferies S, French N, Gilkison C, Graham G, Hope V, Marshall J, et al. COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study. Lancet Public Health. 2020;5(11):e612–23.
49. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science. 2020;368(6492):742–6.
50. Collins OC, Duffy KJ. Estimating the impact of lock-down, quarantine and sensitization in a COVID-19 outbreak: lessons from the COVID-19 outbreak in China. PeerJ. 2020;8: e9933.
51. Braithwaite J, Tran Y, Ellis LA, Westbrook J. The 40 health systems, COVID-19 (40HS, C-19) study. Int J Qual Health Care. 2020;33(1):mzaa113.

52. Koh WC, Naing L, Wong J. Estimating the impact of physical distancing measures in containing COVID-19: an empirical analysis. Int J Infect Dis. 2020;100:42–9.
53. Gupta M, Mohanta SS, Rao A, Parameswaran GG, Agarwal M, Arora M, et al. Transmission dynamics of the COVID-19 epidemic in India and modeling optimal lockdown exit strategies. Int J Infect Dis. 2021;103:579–89.
54. New York Times. See Reopening Plans and Mask Manadates for All 50 States; 2021. https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html.
55. McGrail DJ, Dai J, McAndrews KM, Kalluri R. Enacting national social distancing policies corresponds with dramatic reduction in COVID19 infection rates. PLoS ONE. 2020;15(7): e0236619.
56. Scarabel F, Pellis L, Bragazzi NL, Wu J. Canada needs to rapidly escalate public health interventions for its COVID-19 mitigation strategies. Infect Dis Modell. 2020;5:316–22.
57. Guirao A. The COVID-19 outbreak in Spain, a simple dynamics model, some lessons, and a theoretical framework for control response. Infect Dis Model. 2020;5:652–69.
58. Krishna MV. Mathematical modelling on diffusion and control of COVID-19. Infect Dis Model. 2020;5:588–97.
59. Zhang B, Zhou H, Zhou F. Study on SARS-CoV-2 transmission and the effects of control measures in China. PLoS ONE. 2020;15(11): e0242649.
60. Sebastiani G, Massa M, Riboli E. COVID-19 epidemic in Italy: evolution, projections and impact of government measures. Eur J Epidemiol. 2020;35(4):341–5.
61. Valencia M, Becerra JE, Reyes JC, Castro KG. Assessment of early mitigation measures against COVID-19 in Puerto Rico: March 15–May 15, 2020. PLoS ONE. 2020;15(10): e0240013.
62. Riccardo F, Ajelli M, Andrianou XD, Bella A, Manso MD, Fabiani M, et al. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. Eurosurveillance. 2020;25(49):2000790.
63. Gao S, Rao J, Kang Y, Liang Y, Kruse J, Dopfer D, et al. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. JAMA Netw Open. 2020;3(9): e2020485.
64. Lurie MN, Silva J, Yorlets RR, Tao J, Chan PA. Coronavirus disease 2019 epidemic doubling time in the United States before and during stay-at-home restrictions. J Infect Dis. 2020;222(10):1601–6.
65. Jarvis CI, Zandvoort aKV, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med. 2020;18:124.
66. Ng Y, Li Z, Chua YX, Chaw WL, Zhao Z, Er B, et al. Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore—January 2–February 29, 2020. Morb Mortal Wkly Rep. 2020;69(11):307–11.
67. Cobb JS, Seale MA. Examining the effect of social distancing on the compound growth rate of COVID-19 at the county level (United States) using statistical analyses and a random forest machine learning model. Public Health. 2020;185:27–9.
68. Jardine R, Wright J, Samad Z, Bhutta ZA. Analysis of COVID-19 Burden, epidemiology and mitigation strategies in Muslim majority countries. East Mediter Health J. 2020;26(10):1173–83.
69. Murillo-Zamora E, Guzmán-Esquivel J, Sánchez-Piña RA, Cedeño-Laurent G, Delgado-Enciso I, Mendoza-Cano O. Physical distancing reduced the incidence of influenza and supports a favorable impact on SARS-CoV-2 spread in Mexico. J Infect Develop Countries. 2020;14(9):953–6.
70. Verma BK, Verma M, Verma VK, Abdullah RB, Nath DC, Khan HTA, et al. Global lockdown: an effective safeguard in responding to the threat of COVID-19. J Eval Clin Pract. 2020;26(6):1592–8.
71. Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. BMJ. 2020;370: m2743.
72. Wagner AB, Hill EL, Ryan SE, Sun Z, Deng G, Bhadane S, et al. Social distancing merely stabilized COVID-19 in the United States. Stat. 2020;9(1): e302.
73. Silva L, Filho DF, Fernandes A. The effect of lockdown on the COVID-19 epidemic in Brazil: evidence from an interrupted time series design. Cad Saude Publica. 2020;36(10): e00213920.
74. Medline A, Hayes L, Valdez K, Hayashi A, Vahedi F, Capell W, et al. Evaluating the impact of stay-at-home orders on the time to reach the peak burden of COVID-19 cases and deaths: does timing matter? BMC Public Health. 2020;20:1750.
75. Arshed N, Meo MS, Farooq F. Empirical assessment of government policies and flattening of the COVID 19 curve. J Public Aff. 2020;20: e2333.
76. Holtz D, Zhao M, Benzell SG, Cao CY, Rahimian MA, Yang J, et al. Interdependence and the cost of uncoordinated responses to COVID-19. Proc Natl Acad Sci. 2020;117(33):19837–43.
77. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. Lancet Infect Dis. 2021;21(2):193–202.
78. Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. Science. 2020;369(6500):208–11.
79. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proc Natl Acad Sci. 2020;117(19):10484–91.
80. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369(6500):eabb9789.
81. Anderson SC, Edwards AM, Yerlanov M, Mulberry N, Stockdale JE, Iyaniwura SA, et al. Quantifying the impact of COVID-19 control measures using a bayesian model of physical distancing. PLoS Comput Biol. 2020;16(12): e1008274.
82. Wang T, Wu Y, Lau JYN, Yu Y, Liu L, Li J, et al. A four-compartment model for the COVID-19 infection—implications on infection kinetics, control measures, and lockdown exit strategies. Precis Clin Med. 2020;3(2):104–12.
83. McCarthy Z, Xiao Y, Scarabel F, Tang B, Bragazzi NL, Nah K, et al. Quantifying the shift in social contact patterns in response to non-pharmaceutical interventions. J Math Ind. 2020;10:28.
84. Dandekar R, Rackauckas C, Barbastathis G. A Machine Learning-Aided Global Diagnostic and Comparative Tool to Assess Effect of Quarantine Control in COVID-19 Spread. Patterns. 2020;1(9): 100145.
85. Crokidakis N. COVID-19 Spreading in Rio de Janeiro, Brazil: Do the Policies of Social Isolation Really Work? Chaos, Solitons & Fractals. 2020;136: 109930.
86. Ge J, He D, Lin Z, Zhu H, Zhuang Z. Fourier response system and spatial propagation of COVID-19 in China by a network model. Math Biosci. 2020;330: 108484.
87. Manevski D, Gorenjec NR, Kejžar N, Blagus R. Modeling COVID-19 pandemic using bayesian analysis with application to slovene data. Math Biosci. 2020;329: 108466.

88. Li BZ, Cao NW, Zhou HY, Chu XJ, Ye DQ. Strong policies control the spread of COVID-19 in China. J Med Virol. 2020;92(10):1980–7.
89. Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. Quant Biol. 2020;11–9.
90. Shi Q, Hu Y, Peng B, Tang XJ, Wang W, Su K, et al. Effective control of SARS-CoV-2 transmission in Wanzhou, China. Nat Med. 2020;27:86–93.
91. Adekunle A, Meehan M, Rojas-Alvarez D, Trauer J, McBryde E. Delaying the COVID-19 epidemic in Australia: evaluating the effectiveness of international travel bans. Aust N Z J Public Health. 2020;44(4):257–9.
92. Li Y, Wang LW, Peng ZH, Shen HB. Basic reproduction number and predicted trends of coronavirus disease 2019 epidemic in the Mainland of China. Infect Dis Poverty. 2020;9:94.
93. Kendall M, Milsom L, Abeler-Dörner L, Wymant C, Ferretti L, Briers M, et al. Epidemiological changes on the isle of wight after the launch of the NHS test and trace programme: a preliminary analysis. Lancet Digit Health. 2020;2(12):e658–66.
94. Kang N, Kim B. The Effects of Border Shutdowns on the Spread of COVID-19. J Prev Med Public Health. 2020;53(5):293–301.
95. Tian T, Luo W, Tan J, Jiang Y, Chen M, Pan W, et al. The timing and effectiveness of implementing mild interventions of COVID-19 in large industrial regions via a synthetic control method. Stat Interface. 2021;14(1):3–12.
96. Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early Covid-19 pandemic in the U.S. J Econ. 2021;220(1):23–62.
97. Friedson AI, McNichols D, Sabia JJ, Dave D. Shelter-in-place orders and public health: evidence from California during the COVID-19 pandemic. J Policy Anal Manag. 2020;40(1):258–83.
98. Marschner IC. Back-projection of COVID-19 diagnosis counts to assess infection incidence and control measures: analysis of Australian data. Epidemiol Infect. 2020;148: e97.
99. Valcarcel B, Avilez JL, Torres-Roman JS, Poterico JA, Bazalar-Palacios J, Vecchia CL. The Effect of Early-Stage Public Health Policies in the Transmission of COVID-19 for South American Countries. Rev Panam Salud Publica. 2020;44: e148.
100. Wong CKH, Wong JYH, Tang EHM, Au CH, Lau KTK, Wai AKC. Impact of national containment measures on decelerating the increase in daily new cases of COVID-19 in 54 countries and 4 epicenters of the pandemic: comparative observational study. J Med Internet Res. 2020;22(7): e19904.
101. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. Resurgence of SARS-CoV-2: detection by community viral surveillance. Science. 2021;372(6545):990–5.
102. Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. Nat Commun. 2021;12:1090.
103. Lison A, Persson J, Banholzer N, Feuerriegel S. Estimating the effect of mobility on SARS-CoV-2 transmission during the first and second wave of the COVID-19 epidemic, Switzerland, March to December 2020. Eurosurveillance. 2022;27(10):2100374.
104. Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. Sci Rep. 2020;10:21885.
105. Persson J, Parie JF, Feuerriegel S. Monitoring the COVID-19 epidemic with nationwide telecommunication data. Proc Natl Acad Sci. 2021;118(26): e2100664118.
106. Banholzer N, Feuerriegel S, Vach W (2022) Estimating and explaining cross-country variation in the effectiveness of non-pharmaceutical interventions during COVID-19. Sci Rep 12(1):7526. https://doi.org/10.1038/s41598-022-11362-x
107. Allcott H, Boxell L, Conway JC, Ferguson BA, Gentzkow M, Goldman B. What explains temporal and geographic variation in the early US coronavirus pandemic? National Bureau of Economic Research; 2020. p. 27965.
108. Yan Y, Malik AA, Bayham J, Fenichel EP, Couzens C, Omer SB. Measuring voluntary and policy-induced social distancing behavior during the COVID-19 pandemic. Proc Natl Acad Sci. 2021;118(16): e2008814118.
109. Grossman G, Kim S, Rexer JM, Thirumurthy H. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. Proc Natl Acad Sci. 2020;117(39):24144–53.
110. Herby J. A first literature review: lockdowns only had a small effect on COVID-19. Social Science Research Network; 2021. 3764553.
111. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Commun Health. 2004;58(4):265–71.
112. Bradford A. Association or causation. Proc R Soc Med. 1965;58:295–300.
113. Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. Emerg Themes Epidemiol. 2005;2:11.
114. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol. 2015;12:14.
115. Cox LA. Modernizing the Bradford Hill Criteria for Assessing Causal Relationships in Observational Data. Crit Rev Toxicol. 2018;48(8):682–712.
116. Garin M, Limnios M, Nicolaï A, Bargiotas I, Boulant O, Chick S, Models epidemic, for COVID-19 during the first wave from February to, et al. A methodological review. ArXiv [Preprint]. May 2020;2021(2109):01450.